



## King's Research Portal

DOI:

[10.1016/j.fsigen.2018.02.018](https://doi.org/10.1016/j.fsigen.2018.02.018)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Aliferi, A., Thomson, J., McDonald, A., Paynter, V. M., Ferguson, S., Vanhinsbergh, D., Court, D. S., & Ballard, D. (2018). UK and Irish Y-STR population data—A catalogue of variant alleles. *Forensic Science International-Genetics*, 34, E1-E6. <https://doi.org/10.1016/j.fsigen.2018.02.018>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

## Accepted Manuscript

Title: UK and Irish Y-STR population data—A catalogue of variant alleles

Authors: Anastasia Aliferi, Jim Thomson, Andrew McDonald, Vanessa Molin Paynter, Steven Ferguson, Des Vanhinsbergh, Denise Syndercombe Court, David Ballard



PII: S1872-4973(18)30008-5  
DOI: <https://doi.org/10.1016/j.fsigen.2018.02.018>  
Reference: FSIGEN 1859

To appear in: *Forensic Science International: Genetics*

Received date: 5-1-2018  
Revised date: 13-2-2018  
Accepted date: 22-2-2018

Please cite this article as: Anastasia Aliferi, Jim Thomson, Andrew McDonald, Vanessa Molin Paynter, Steven Ferguson, Des Vanhinsbergh, Denise Syndercombe Court, David Ballard, UK and Irish Y-STR population data—A catalogue of variant alleles, *Forensic Science International: Genetics* <https://doi.org/10.1016/j.fsigen.2018.02.018>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Short communication

## UK and Irish Y-STR population data – a catalogue of variant alleles

Anastasia Aliferi<sup>a</sup>, Jim Thomson<sup>b</sup>, Andrew McDonald<sup>c</sup>, Vanessa Molin Paynter<sup>a</sup>, Steven Ferguson<sup>d</sup>, Des Vanhinsbergh<sup>e</sup>, Denise Syndercombe Court<sup>a</sup>, David Ballard<sup>a\*</sup>

<sup>a</sup>*King's Forensics, Faculty of Life Sciences and Medicine, King's College London, UK*

<sup>b</sup>*Eurofins Forensic Services, Teddington, UK*

<sup>c</sup>*Cellmark Forensic Services, Abingdon, UK*

<sup>d</sup>*Scottish Police Authority Forensic Services, Edinburgh, UK*

<sup>e</sup>*Key Forensic Services, Wymondham, Norfolk, UK*

\* Corresponding author at: Department of Analytical Environmental and Forensic Sciences, King's College London (Waterloo Campus, Franklin Wilkins Building), 150 Stamford Street, London SE1 9NH

E-mail address: [david.ballard@kcl.ac.uk](mailto:david.ballard@kcl.ac.uk) (D. Ballard)

## Highlights

- Reporting 3128 Powerplex Y23 profiles from four UK and Irish populations
- Investigation into shared haplotypes using 5 additional loci included in Yfiler Plus kit
- Investigation into allele duplications and deletions
- Sequencing of primer binding and mobility shift mutations in loci DYS481, DYS392 and DYS576

## Abstract

A total of 3128 Y-STR profiles from three UK and one Irish population have been analysed with the PowerPlex Y23 system and are reported here. Instances of haplotype sharing between apparently unrelated individuals were identified and further investigated with the use of the 5 additional markers within the Yfiler Plus kit, resulting in a reduction by 76% in the number of shared haplotypes. Furthermore, Yfiler Plus was also employed to verify locus deletions and duplications observed in Y23 genotypes while inconsistencies between the two kits were sequenced, revealing underlying Y23 primer binding site mutations in loci DYS392 and DYS576. Finally, the mechanism behind a previously reported population specific peak shift observed in DYS481 in South Asian samples has been evaluated and further investigated in a novel case of this phenomenon seen in a Black British individual featuring a different flanking region mutation.

## 1. Introduction

Match frequency estimation of Y-STR haplotypes between unrelated individuals is vital when assessing the weight of evidence in cases using Y chromosome STR markers. By the start of 2017, the Y Haplotype Reference Database (YHRD) contained only a few hundred haplotypes from the UK and Ireland that consisted of at least 22 Y-STR markers [1], this being the minimum number of markers routinely used in criminal casework, and the low number of haplotypes was identified as an issue at the national level. The UK YHRD expansion project was initiated to address this need and resulted in the generation of 3128 Y-STR profiles consisting of at least 22 markers, that are detailed here. The collection of over 3000 haplotypes from four distinct population groups provided an opportunity to both thoroughly investigate the variant allele patterns that can be observed in a study of this size and to assess the efficacy of a 22 versus 27 Y-STR marker strategy with respect to individualisation of haplotypes.

## 2. Materials and methods

### 2.1 Population

Samples in this study consisted of buccal swabs and blood samples collected from 3128 residents of the United Kingdom (UK) or Ireland. The majority of the samples were analysed at the DNA Analysis at King's facility at King's College London, while the UK Association of Forensic Science Providers (AFSP) provided profiles for 334 samples. Informed consent was obtained for all cases and relationships between the donors were checked to ensure that all individuals were unrelated.

Samples were divided into four populations based on self-declared ethnicity of the donors: White British (UK-EA1), Irish (IRE-EA1), Black British (UK-EA3) and British South Asian (UK-EA4). The White British population included 1062 samples from donors that had declared their ethnicity as British, White or Caucasian. Samples obtained from donors that were self-declared as Irish comprised the Irish population in this study and were 720 in total. The Black British population included 977 samples from individuals with ancestral origins from African or Caribbean countries; if stated these were principally Nigeria, Somalia and Ghana. Finally, the British South Asian population comprised of 369 samples collected from donors with ancestral origins from Bangladesh, India, Pakistan, Sri Lanka and Afghanistan. Additionally, 526 profiles from the same populations, previously submitted to the YHRD database by the authors [1], were included in the results analysis when assessing variant allele frequencies.

### 2.2 DNA extraction

DNA was isolated from buccal swabs or whole blood using the Chelex extraction method [2] and was stored at 4 °C.

### 2.3 DNA amplification

Twenty-two markers containing short tandem repeats (STRs) located in 23 positions on the Y chromosome were targeted using the PowerPlex Y23 System (Promega). The analysed STRs include: DYS19, DYS385 (a multicopy marker), DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS643, DYS635 and GATA-H4. All markers were analysed in one multiplex PCR reaction according to the manufacturer's guidelines, validated to half reaction volumes. Individuals presenting abnormalities in their Y23 profiles (e.g. marker deletions/duplications) were re-analysed using the Yfiler Plus PCR Amplification Kit (Thermo-Fisher Scientific), which incorporates the majority of markers included in Powerplex Y23 (all except DYS549 and DYS643) with the addition of five extra Y-STRs: DYS387-S1,

DYS449, DYS460, DYS518 and DYS627. In cases where the original DNA extract was still available, the same kit was also used to further investigate samples exhibiting identical Y23 profiles. The Yfiler Plus kit was used according to the manufacturer's guidelines, employing half volumes. Finally, new primers were designed for DYS481, DYS392 and DYS576 (Supplementary Table S1) in order to sequence the loci where inconsistencies between the two kits were observed. PCR was performed in 10  $\mu$ L reactions using the Qiagen Multiplex PCR Master Mix with final primer concentrations of 0.3  $\mu$ M and the addition of 1 ng of DNA. Amplification conditions used for these loci were as follows: 95 °C for 15 min followed by 30 cycles of 95 °C for 15 s, 59 °C for 15 s and 72 °C for 30 s. Products were treated with 1  $\mu$ L of ExoStar 1-Stop and incubated at 37°C for 60 min followed by 15 min at 80°C. Finally, sequencing was performed using BigDye Terminator v3.1 (Applied Biosystems) according to manufacturer's recommendations.

## 2.4 Detection and genotyping

Following amplification, electrophoresis of PCR products was carried out using a 3130xl genetic analyser (Applied Biosystems) in combination with POP-7 polymer. For STR genotyping, allelic ladders were provided with the kits and run in parallel with the samples.

## 2.5 Analysis of data

STR profile analysis was conducted using GeneMapper ID-X analysis software (Thermo-Fisher Scientific). Sequencing analysis to detect single nucleotide polymorphisms (SNPs) was performed using SeqScape (Applied Biosystems).

## 2.6 Predictions of secondary structures

The Mfold software (<http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form>) [3,4] was used for the prediction of secondary structure formation of a DNA sequence regarding locus DYS481.

## 2.7 Quality control

Proficiency is externally assessed by the College of American Pathologists and/or GEDNAP. Certification from YHRD (<http://www.yhrd.org>) has been granted since 2001. This paper follows the guidelines for publication of population genetics data [5-6] and the data are available on YHRD with accession numbers YA003115, YA003116, YA003117 and YA003118.

# 3. Results and discussion

The observed haplotypes can be accessed through Supplementary Table S2 for the White British (UK-EA1), Irish (IRE-EA1), Black British (UK-EA3) and British South Asian (UK-EA4) populations.

## 3.1 Haplotype Frequencies

The majority of the haplotypes recorded in this study occur only once, providing a total of 3062 different Y23 haplotypes from the 3128 individuals, however certain haplotypes were shared within populations. Analysis with Powerplex Y23 revealed 5/1057 haplotypes appearing twice in the White British population (UK-EA1), 13/704 haplotypes appearing twice and 2/704 haplotypes appearing 3 times in the Irish population (IRE-EA1), 11/954 haplotypes appearing twice, 2/954 haplotypes appearing 3 times and 1/954 haplotypes appearing 8 times in the Black British population (UK-EA3) and finally, 7/358 haplotypes appearing twice and 1/358 haplotypes appearing 5 times in the British South Asian population (UK-EA4) (Table 1A). Additionally, 11 haplotypes appear both in the White British and in the Irish populations, 3 of which already appear twice in the Irish population and 1 of which appears three times in the Irish population.

Where possible these samples were further analysed with the Yfiler Plus kit. The addition of five extra STRs increased the diversity significantly, raising the number of different haplotypes recorded in this study to 3111, and helped separate a large percentage of the previously shared haplotypes. With the additional markers, the number of haplotypes appearing twice reduced in all populations, from 13/704 to 1/719 in the Irish population, from 11/954 to 4/969 in the Black British population and from 7/358 to 2/363 in the British South Asian population – in the White British population it was only possible to genotype 4 out of 5 sets of shared haplotypes, however within these 4 sets of samples, the addition of the extra Yfiler Plus markers was sufficient to reduce the number of haplotypes shared from 4/954 to 1. The number of haplotypes appearing 3 times was reduced to 0 for all four populations, and the haplotype that appeared 8 times in the Black British population was separated to a haplotype appearing 4 times and a haplotype appearing twice, while the haplotypes of two of the initial 8 samples no longer matched any of the rest. Finally, the frequency of the haplotype appearing 5 times in the British South Asian population did not change even after the addition of the extra markers (Table 1B). Further investigation of the ancestry of the 4 Black British and 5 British South Asian individuals that continued to exhibit the same haplotype within the population group even after analysis with the Yfiler Plus chemistry, revealed that they belonged to individuals originating from Somalia and Afghanistan respectively. These populations are known to show high levels of stratification that partly explains these observations [7-9].

### 3.2 Chromosome Structure Variations

#### 3.2.1 Duplications

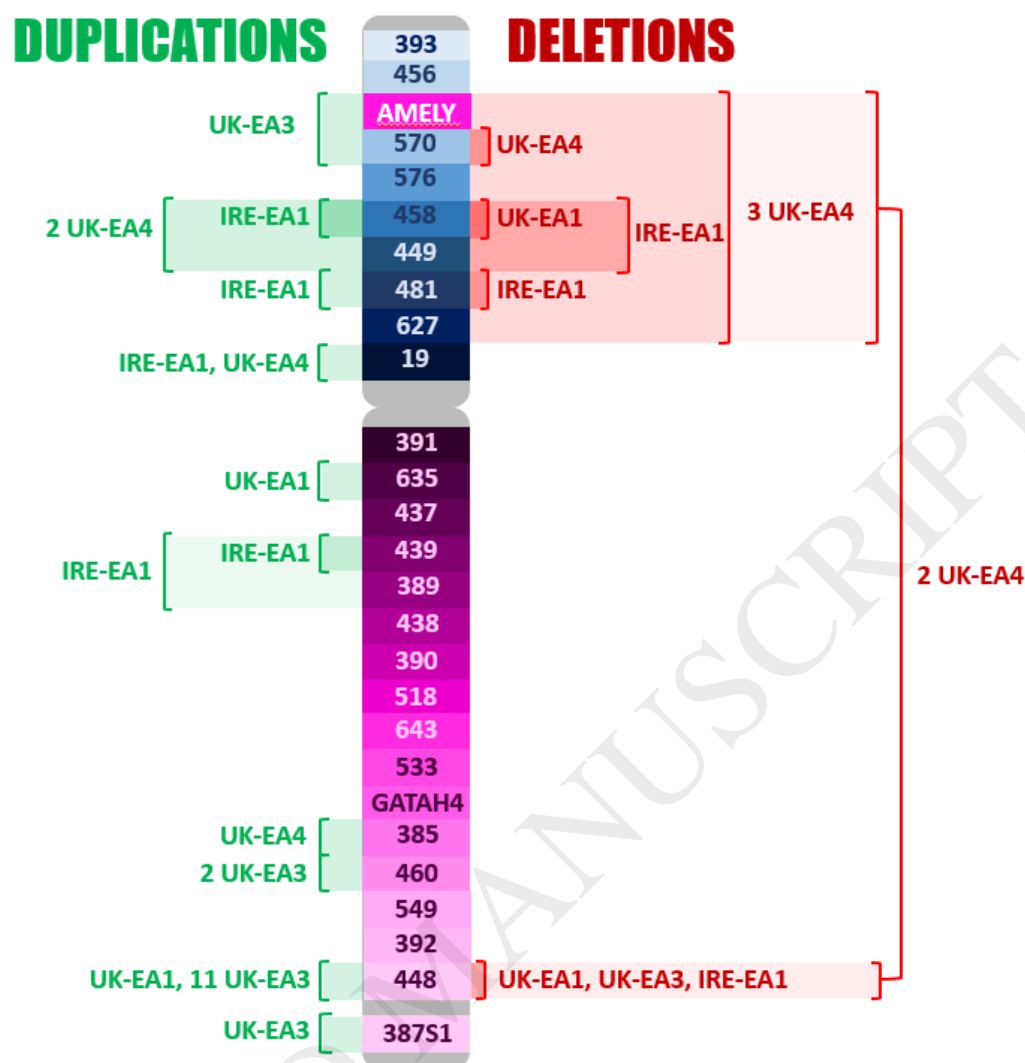
Analysis with Powerplex Y23 and Yfiler Plus revealed duplications (two allele peaks in a single locus) of 11 loci. Adjacent loci DYS458 and DYS449 were duplicated simultaneously in 2 British South Asian individuals (UK-EA4), DYS570 was duplicated in one Black British individual (UK-EA3), DYS481 was duplicated in one Irish individual (IRE-EA1), DYS19 was duplicated in one British South Asian individual (UK-EA4) and one Irish individual (IRE-EA1), DYS635 was duplicated in one White British individual (UK-EA1), DYS439 was duplicated in one Irish individual (IRE-EA1) and in a second individual of the same population DYS439 was duplicated along with the adjacent DYS389. Finally, a duplication of DYS387-S1 was observed in one Black British individual (UK-EA3) and a duplication of DYS448 was recorded in one White British (UK-EA1) and 7 Black British individuals (UK-EA3) (frequency of 0.0072). The duplication in DYS448 has been previously recorded in 4 different Black British individuals [1] suggesting a population specific phenomenon appearing in approximately 1% (0.0095) of the Black British individuals. These results are summarised in Table 2 and a visual representation can be found in Figure 1.

#### 3.2.2 Deletions

STR deletions were confirmed in 8 individuals and included 8 loci. An extended deletion around the Amelogenin region including DYS570, DYS576, DYS458, DYS449, DYS481 and DYS627 was observed in one British South Asian individual (UK-EA4). The same deletion has been previously observed in 2 different individuals from the same British South Asian population [1] giving a final frequency of 0.0059 for this extensive Y chromosome deletion in the population. In another individual from the same population, the extensive deletion described above was observed but additionally included the deletion of locus DYS448, located on the opposite end of the chromosome. This simultaneous deletion of loci DYS570, DYS576, DYS458, DYS449, DYS481, DYS627 and DYS448 has been previously recorded in one other British South Asian individual (final frequency of 0.0039). Large deletions encompassing the AMELY region and involving the adjacent loci DYS570, DYS576, DYS458, DYS481 as well as the distant locus DYS448 have been previously reported in individuals with South Asian ancestry [1,10-

11]. Loci DYS449 and DYS627, included in the Yfiler Plus PCR kit, were not analysed in these previous studies. However, given their close proximity to the other loci involved in this extensive deletion it is of no surprise to find these loci additionally deleted when genotyping with Yfiler Plus, and indeed the typing of DYS627 further identifies the bounds of this deletion. A single locus deletion of DYS448 was also observed in one White British (UK-EA1), one Black British (UK-EA3), and one Irish individual (IRE-EA1). Finally, deletion of DYS458 and DYS449 was observed in one Irish individual (IRE-EA1), while a deletion of DYS458 alone was observed in a White British individual (UK-EA1) and a single deletion of DYS481 was detected in an Irish individual (IRE-EA1). An interesting observation is that the Amelogenin locus was deleted in 4 out of 5 individuals with a DYS570 deletion and was duplicated in the 1 individual with a DYS570 duplication, suggesting that Amelogenin is closely linked to locus DYS570 due to the close proximity of the two loci. A summary of these observations can be found in Table 3 and a visual representation can be found in Figure 1.

Overall it is worth noting the distinctive areas highlighted in Figure 1, suggesting increased instability in specific regions of the Y chromosome: 17 out of 37 observed chromosome structural variants included DYS448 while 10 of the 37 included DYS458, and both of these examples involve multiple different forms of structural variation in diverse populations, implying independent origins for many of these changes. Previous studies have highlighted how highly similar (e.g. palindromic) sequences on the Y chromosome can facilitate gene conversion in specific chromosomal regions, leading to an increase in independent deletion and mutation events at certain loci, including at DYS448 [12]. Conversely, some specific structural changes, for example the extended deletions around the Amelogenin region, were observed multiple times in only one population indicating a single origin for that specific variant that has come to prominence through genetic drift.



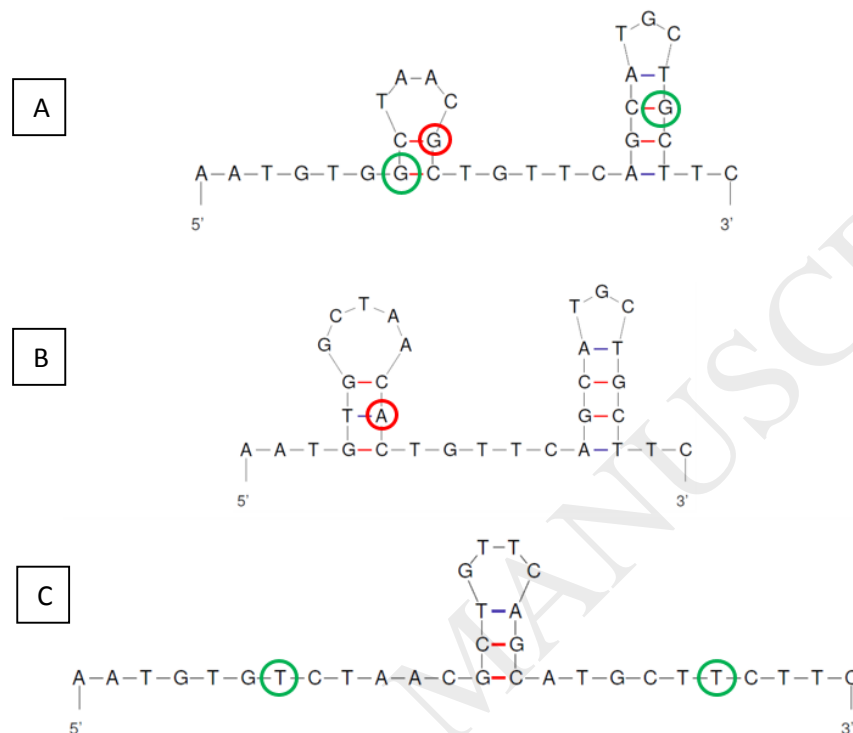
**Figure 1.** Visual representation of the observed duplications and deletions in the merged dataset of 3654 individuals, including 526 individuals reported by Purps *et al.* [1].

### 3.3 Mutations

A mobility shift in the capillary electrophoresis peak, causing it to be falsely reported as an off-ladder allele, was observed in 27 out of 369 British South Asian individuals (frequency of 0.0732) for locus DYS481, associated predominantly with allele 25. Sequencing revealed an underlying single nucleotide polymorphism from G>A (rs368663163) 16bp upstream of the CTT repeat region of the locus. This SNP appears in the R2 haplogroup and has been previously reported in South Asian populations with a similar frequency [13]. Lee *et al.* in their study reported that the mutation changes a small secondary structure, which is naturally formed in the amplicon around this nucleotide region, to a larger stem-loop structure, thus resulting in the observed change in electrophoretic mobility (Figure 2) [13]. This mutation appears to be population specific and can potentially be used for South Asian ancestry inference given its relatively high frequency within the population. In a novel case, the same peak shift in DYS481 was also observed in a single Black British individual in this study (frequency 0.001). However, sequencing of this sample revealed a different flanking region change - two G>T mutations located at 1 (rs370750300) and 22 (genome assembly GRCh38 coordinate Y:8,558,315) bases upstream of the DYS481 repeat region. Simulation with the Mfold software suggests that this double



mutation alters the secondary structure formed in this region, disrupting the 2 original loops and resulting in a different structure (Figure 2). These results, together with the observations from Lee *et al.* [13], suggest that any mutation altering the loop formation in this region could potentially result in a mobility peak shift in capillary electrophoresis giving a false off-ladder call, and further strengthens the evidence that certain flanking region changes result in electrophoretic mobility shifts even in the presence of a denaturing electrophoretic environment. For this reason, extra care is advised when reporting results involving DYS481.



**Figure 2.** Mutations observed upstream of the DYS481 locus resulting in shifted electropherogram peaks following capillary electrophoresis. Two variations from the reference (A) sequence were observed, a single G>A transition was observed in 27 British South Asian individuals (UK-EA4) (B) and two G>T transversions were recorded in an individual from the Black British population (UK-EA3) (C). In the first case the first loop of the secondary structure appears to extend, while the mutations in the second case disrupt both loops and result in the formation of a less extensive secondary structure.

Single nucleotide polymorphisms resulting in null alleles were observed in 2 loci: DYS392 and DYS576 (Figure 3). A change from C>T was recorded 258bp upstream of the TAT repeat region of DYS392 (genome assembly GRCh38 coordinate Y:20,472,283) in 3 Black British individuals (frequency of 0.0031) all presenting with null DYS392 alleles. A similar change from C>T was observed 19bp downstream of the AAAG repeat region of DYS576 in 4 British South Asian (0.0108) and 1 Black British individual (0.0010) exhibiting a null allele at DYS576. This DYS576 null allele has also been observed in 2 British South Asian individuals previously reported by this group [1], giving a British South Asian frequency for the null allele of 1.2% (0.0118) in the merged dataset. Both individuals from this previous study also carried this C>T mutation which is a known SNP (rs757752030) with a reported frequency of 1.4% in the South Asian 1000 genomes data [14]. These results suggest that

extra care should be taken when analysing DYS576 in this population given the relatively high occurrence of this primer-binding site mutation.

## DYS392

3 UK-EA3 samples

TAATTTGGGGTAGTGCTTAGTCCCCTCTTGGCATCTAGTGA<sup>T</sup>CTCATATCAGATAGAGGCAGTCATCGCAGTGG  
 CCCAAGTGATCTTGCAACATCTCCATCCATGTTGCTCCAAAGGACCCAATTTACTGTAAATGGTTGTATAGTATT  
 TATGGTCTACATAGACCATATTTACCATATGTTTCATCCATATTTCTTCATTAATCTAGCTTTTAAAAACAATAATTTG  
 ATTTCAAGTGTGTTATTTAAAAGCCAAGAAGGAAAACAAATTTTTCTGTATCACCATTATTTATTATTATTATTA  
 TTATTATTATTATTATTATTATTATTATTACTAAGGAATGGGATTGGTAGGTTAATGATCCCTCTGTTTTGACTTCT  
 TTGAGATATTTCCAGACTACTTTCCACTTTGACTGTAGGAATTTACATTGCATCAACTGGGTCTAAGTATTCTTTTT  
 TTCTGAAGTCTTGCTAGCTTCCATAATTTCA

## DYS576

7 UK-EA4 and 1 UK-EA3 sample

TTCAATCTCAGCCAAGCAACATAGCAAGACCTCATCTCTGAATAAAAGAAAGAAAGAAAGAAAGAAAGAAAG  
 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG  
 AAAAGCCAAGACAAATACGCTTATTACTC  
 CCATCTCCTCCTTCATCTCCAGGAAATGAGACTGCCATACTGATATGGT

**Figure 3.** Mutations observed in loci DYS392 and DYS576 in individuals where multiplex-specific null alleles were observed. Repeat regions are highlighted in blue and mutations in pink.

## 4. Conclusions

Overall, this study reports 3062 different Y23 haplotypes from the four main UK populations. With the addition of the 5 extra Y-STR markers of the Yfiler Plus chemistry, the number of different haplotypes increased to 3111 as the shared haplotypes decreased from 42 to 10. However, even though a decrease of 76% was observed in the shared haplotypes when Yfiler Plus was included in the analysis, the overall increase of different haplotypes only amounted to 1.6%. Additionally, an extensive catalogue of loci deletions and duplications have been identified in this multi-population data set that show consistent patterns of structural variation in hotspots on the Y chromosome.

Discordances between the PowerPlex Y23 and the Yfiler Plus kits highlighted nucleotide polymorphisms surrounding the repeat regions in four loci. These findings include 2 SNPs in the primer binding regions of DYS392 (novel SNP) and DYS576, with the first reported solely in individuals of Black British origins and the second identified predominantly in individuals of South Asian origin. Additionally, a previously reported mobility shift in Y23 generated electrophoretic peaks was observed not only in the British South Asian population at a frequency matching that of previous reports (~7%), but was also observed for the first time in an individual of African origin. Subsequent sequencing revealed the presence of a previously identified SNP in the South Asian individuals upstream of loci DYS481 [13], and identified two novel polymorphisms in the same region in the Black British individual. These findings, together with structure analysis performed with the Mfold software, provide further support for the mechanism proposed by Lee *et al.* [13] for this mobility shift that focuses on the disruption of a specific ampliconic secondary structure formed in this region.

**Acknowledgments**

This work was carried out as part of the UK YHRD expansion project, coordinated by the Association of Forensic Science Providers and funded by Yorkshire and the Humber Policing. The authors would like to thank Kevin Morton, Director of Scientific Services, for his assistance.

## References

- [1] Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., Angustia, S.M.T., Santos, L.H., Anslinger, K., Bayer, B., Ayub, Q., Wei, W., Xue, Y., Tyler-Smith, C., Bafalluy, M.B., Martínez-Jarreta, B., Egyed, B., Balitzki, B., Tschumi, S., Ballard, D., Court, D.S., Barrantes, X., Bäßler, G., Wiest, T., Berger, B., Niederstätter, H., Parson, W., Davis, C., Budowle, B., Burri, H., Borer, U., Koller, C., Carvalho, E.F., Domingues, P.M., Chamoun, W.T., Coble, M.D., Hill, C.R., Corach, D., Caputo, M., D'Amato, M.E., Davison, S., Decorte, R., Larmuseau, M.H.D., Ottoni, C., Rickards, O., Lu, D., Jiang, C., Dobosz, T., Jonkisz, A., Frank, W.E., Furac, I., Gehrig, C., Castella, V., Grskovic, B., Haas, C., Wobst, J., Hadzic, G., Drobnic, K., Honda, K., Hou, Y., Zhou, D., Li, Y., Hu, S., Chen, S., Immel, U., Lessig, R., Jakovski, Z., Ilievska, T., Klann, A.E., García, C.C., de Knijff, P., Kraaijenbrink, T., Kondili, A., Miniati, P., Vouropoulou, M., Kovacevic, L., Marjanovic, D., Lindner, I., Mansour, I., Al-Azem, M., Andari, A.E., Marino, M., Furfuro, S., Locarno, L., Martín, P., Luque, G.M., Alonso, A., Miranda, L.S., Moreira, H., Mizuno, N., Iwashima, Y., Neto, R.S.M., Nogueira, T.L.S., Silva, R., Nastainczyk-Wulf, M., Edelmann, J., Kohl, M., Nie, S., Wang, X., Cheng, B., Núñez, C., Pancorbo, M.M.d., Olofsson, J.K., Morling, N., Onofri, V., Tagliabracci, A., Pamjav, H., Volgyi, A., Barany, G., Pawlowski, R., Maciejewska, A., Pelotti, S., Pepinski, W., Abreu-Glowacka, M., Phillips, C., Cárdenas, J., Rey-Gonzalez, D., Salas, A., Brisighelli, F., Capelli, C., Toscanini, U., Piccinini, A., Piglionica, M., Baldassarra, S.L., Ploski, R., Konarzewska, M., Jastrzebska, E., Robino, C., Sajantila, A., Palo, J.U., Guevara, E., Salvador, J., Ungria, M.C.D., Rodriguez, J.J.R., Schmidt, U., Schlauderer, N., Saukko, P., Schneider, P.M., Sirker, M., Shin, K., Oh, Y.N., Skitsa, I., Ampati, A., Smith, T., Calvit, L.S.d., Stenzl, V., Capal, T., Tillmar, A., Nilsson, H., Turrina, S., De Leo, D., Verzeletti, A., Cortellini, V., Wetton, J.H., Gwynne, G.M., Jobling, M.A., Whittle, M.R., Sumita, D.R., Wolańska-Nowak, P., Yong, R.Y.Y., Krawczak, M., Nothnagel, M. and Roewer, L. **A global analysis of Y-chromosomal haplotype diversity for 23 STR loci.** *Forensic Sci. Int. Genet.* 2014; 12: 12-23.
- [2] Walsh, P.S., Metzger, D.A. and Higuchi, R. **Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material.** *Biotechniques* 1991; 10: 4: 506-513
- [3] Mathews, D.H. **Predicting RNA secondary structure by free energy minimization.** *Theor. Chem. Acc.* 2005; 116:160-168.
- [4] Zuker, M. **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res.* 2003; 31: 3406-3415
- [5] Carracedo, Á., Butler, J.M., Gusmão, L., Linacre, A., Parson, W., Roewer, L. and Schneider, P.M. **New guidelines for the publication of genetic population data.** *Forensic Sci. Int. Genet.* 2013; 7: 217-220
- [6] Gusmão, L., Butler, J.M., Linacre, A., Parson, W., Roewer, L., Schneider, P.M. and Carracedo, A. **Revised guidelines for the publication of genetic population data.** *Forensic Sci. Int. Genet.* 2017; 30: 160-163
- [7] Hallenberg, C., Simonsen, B., Sanchez, J. and Morling, N. **Y-chromosome STR haplotypes in Somalis.** *Forensic Sci. Int.* 2005; 151: 317-321
- [8] Achakzai, N.M., Rahman, Z., Shahzad, M.S., Daud, S., Zar, M.S., Israr, M., Husnain, T., Willuweit, S. and Roewer, L. **Y-chromosomal STR analysis in the Pashtun population of Southern Afghanistan.** *Forensic Sci. Int. Genet.* 2012; 6: e103-e105

- [9] Tillmar, A.O. and Montelius, K. **Population data of 12 Y-STR loci from a Somali population.** *Forensic Sci. Int. Genet. Supp. Series* 2009; 2: 413-415
- [10] Chang, Y.M., Perumal, R., Keat, P.Y., Yong, R.Y.Y., Kuehn, D.L.C. and Burgoyne, L. **A distinct Y-STR haplotype for Amelogenin negative males characterized by a large Y<sub>p</sub>11.2 (DYS458-MSY1-AMEL-Y) deletion.** *Forensic Sci. Int. Genet.* 2007; 166: 115-120
- [11] Takayama, T., Takada, N., Suzuki, R., Nagaoka, S., Watanabe, Y., Kumagai, R., Aoki, Y. and Butler, J.M. **Determination of deleted regions from Yp11.2 of an amelogenin negative male.** *Leg Med* 2009; 11: 578-580
- [12] Balaesue, P., Bowden, G.R., Parkin, E.J., Omran, G.A., Heyer, E., Quintana-Murci, L., Roewer, L., Stoneking, M., Nasidze, I., Carvalho-Silva, D.R., Tyler-Smith, C., de Knijff, P. and Jobling, M.A. **Dynamic Nature of the Proximal AZFc Region of the Human Y Chromosome: Multiple Independent Deletion and Duplication Events Revealed by Microsatellite Analysis.** *Hum Mutat.* 2008; 29(10): 1171-1180
- [13] Lee, E.Y., Lee, H.Y. and Shin, K. **Off-ladder alleles due to a single nucleotide polymorphism in the flanking region at DYS481 detected by the PowerPlex® Y23 System.** *Forensic Sci. Int. Genet.* 2016; 24: 7-8
- [14] The Genomes Project, C. **A global reference for human genetic variation.** *Nature* 2015; 526:68

## Tables

**Table 1**

Number of individuals sharing the same Y23 (A) and Y23 and YfilerPlus combined haplotype (B) within the different populations.

A	Population	Number of individuals analysed	Number of different Y23 haplotypes	Times a Y23 haplotype is observed			
				x2	x3	x5	x8
	White British (UK-EA1)	1062	1057	5			
	Irish (IRE-EA1)	720	703	13	2		
	Black British (UK-EA3)	977	955	11	2		1
	British South Asian (UK-EA4)	369	358	7		1	

B	Population	Number of individuals analysed	Number of different haplotypes after Yfiler Plus analysis	Times a Y23 and Yfiler Plus haplotype is observed		
				x2	x4	x5
	White British (UK-EA1)	1062	1061	1*		
	Irish (IRE-EA1)	720	719	1		
	Black British (UK-EA3)	977	970	4	1	
	British South Asian (UK-EA4)	369	363	2		1

\*In the White British population only 4/5 haplotypes shared between 2 individuals could be extended with Yfiler Plus, and of those 4 haplotypes, 1 was still identical even after the addition of the extra 5 loci.

**Table 2**

Observed locus duplications in the current dataset and in a merged dataset including 526 profiles from Purps *et al.* previously tested in this laboratory [1].

Duplicated Locus/Loci	Number of individuals in this study (n=3128)	Number of individuals in merged studies (n=3654)	Ancestry (and frequency in the two datasets for each population)
DYS458, DYS449	2	2	UK-EA4 (0.0054, 0.0039)
DYS458		1	IRE-EA1 (0, 0.0013)
DYS570	1	1	UK-EA3 (0.0010, 0.0009)
DYS481	1	1	IRE-EA1 (0.0014, 0.0013)
DYS19	2	2	1 IRE-EA1 (0.0014, 0.0013) and 1 UK-EA4 (0.0027, 0.0020)
DYS635	1	1	UK-EA1 (0.0009, 0.0008)
DYS439	1	1	IRE-EA1 (0.0014, 0.0013)
DYS439, DYS389	1	1	IRE-EA1 (0.0014, 0.0013)
DYS387-S1	1	1	UK-EA3 (0.0010, 0.0009)
DYS448	8	12	1 UK-EA1 (0.0009, 0.0008) and 11 UK-EA3 (0.0072, 0.0095)
DYS385		1	UK-EA4 (0, 0.0020)

<b>DYS460</b>	<b>2</b>	<b>2</b>	<b>UK-EA3 (0.0010, 0.0009)</b>
---------------	----------	----------	--------------------------------

**Table 3**

Observed loci deletions in the current dataset and in a merged dataset including 526 profiles from Purps *et al.* previously tested in this laboratory [1].

<b>Deleted Locus/Loci</b>	<b>Number of individuals in this Study (n=3128)</b>	<b>Number of individuals in merged studies (n=3654)</b>	<b>Ancestry (and frequency in the two datasets for each population)</b>
<b>DYS458</b>	<b>1</b>	<b>1</b>	<b>UK-EA1 (0.0009, 0.0008)</b>
<b>DYS570</b>		<b>1</b>	<b>UK-EA4 (0, 0.0020)</b>
<b>DYS481</b>	<b>1</b>	<b>1</b>	<b>IRE-EA1 (0.0014, 0.0013)</b>
<b>DYS458, DYS449</b>	<b>1</b>	<b>1</b>	<b>IRE-EA1 (0.0014, 0.0013)</b>
<b>DYS570, DYS576, DYS458, DYS449, DYS481, DYS627</b>	<b>1</b>	<b>3</b>	<b>UK-EA4 (0.0027, 0.0059)</b>
<b>DYS570, DYS576, DYS458, DYS449, DYS481, DYS627, DYS448</b>	<b>1</b>	<b>2</b>	<b>UK-EA4 (0.0027, 0.0039)</b>
<b>DYS448</b>	<b>3</b>	<b>3</b>	<b>1 UK-EA1 (0.0009, 0.0008), 1 UK-EA3(0.0010, 0.0009), 1 IRE-EA1 (0.0014, 0.0013)</b>